ESIP Earth Science Data Analytics (ESDA) Cluster

November 20, 2014

Agenda

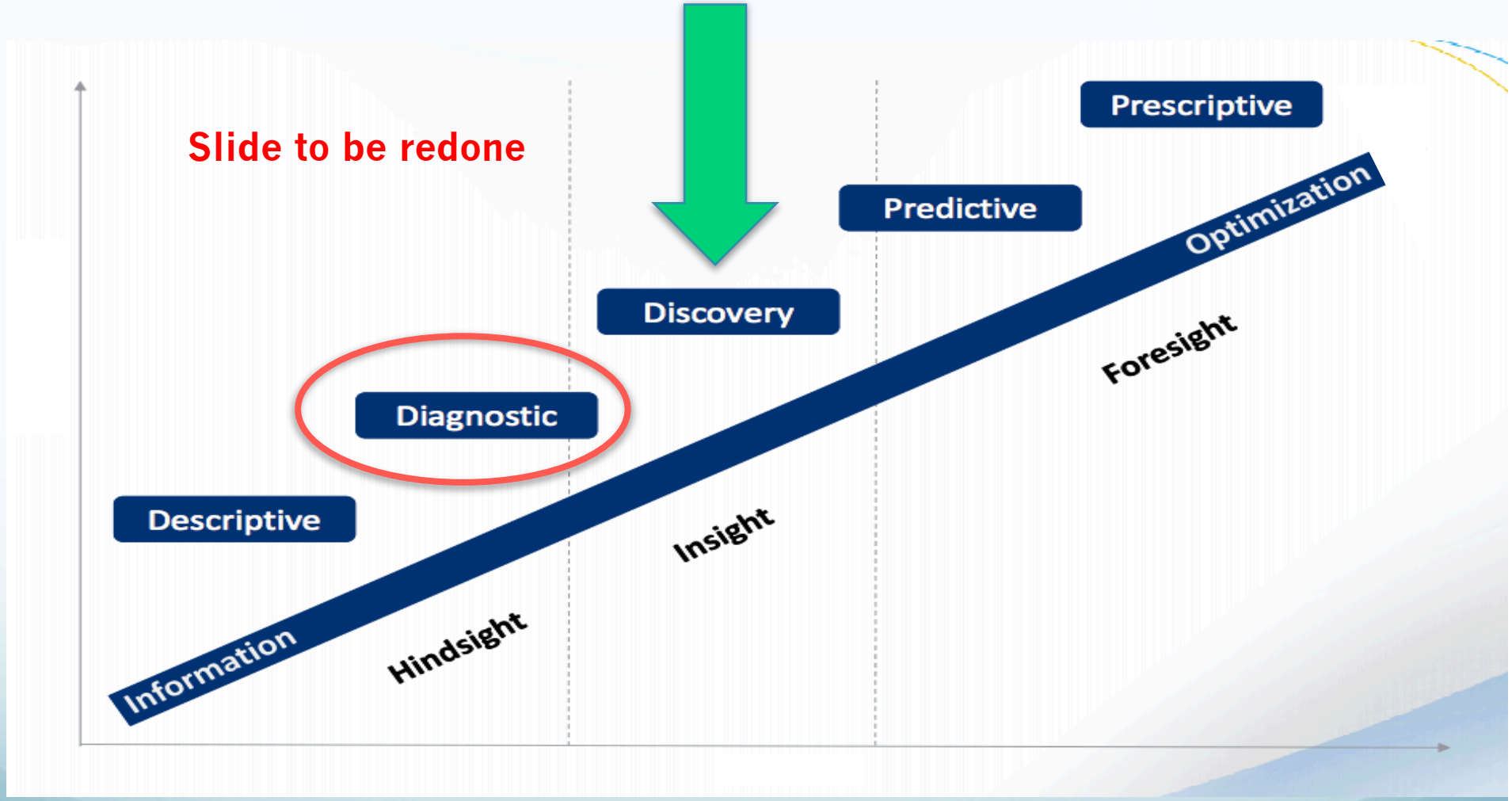
- 10 min – Recap of our last telecon on Diagnostic Analytics
- 20 minutes – Discussion: Descriptive and Predictive Analytics (and expand our current Data Analytics Type Comparisons Table. For current table see: http://wiki.esipfed.org/index.php/Earth_Science_Data_Analytics/2014-10-23_Telecon)
- 20 minutes – Planning ahead discussion:

  - Winter ESIP Meeting ESDA Planning: Sessions; Suggestions for guest speakers
  - Are we starting to learn enough to write a paper on the Types of Data Analytics Utilized in (the various phases of) Earth Science

- 10 minutes - Open Mic – Thoughts, Ideas

Discovery Analytics:

This is where people learn from the data.



http://www.informationbuilders.es/intl/co.uk/presentations/four_types_of_analytics.pdf

Discussion – Diagnostic Analytics

Descriptive Analytics: You can quickly understand "what happened" during a given period in the past and verify if a campaign was successful or not based on simple parameters.

Diagnostic Analytics: If you want to go deeper into the data you have collected from users in order to understand "Why some things happened," you can use … intelligence tools to get some insights.

Discoveritive Analytics: The use of data and analysis tools/models to discover information

Predictive Analytics: If you can collect contextual data and correlate it with other user behavior datasets, as well as expand user data … you enter a whole new area where you can get real insights.

Prescriptive Analytics: Once you get to the point where you can consistently analyze your data to predict what's going to happen, you are very close to being able to understand what you should do in order to maximize good outcomes and also prevent potentially bad outcomes. This is on the edge of innovation today, but it's attainable!

Modified from: http://www.ciandt.com/card/four-types-of-analytics-and-cognition

Discussion – Diagnostic Analytics

Diagnostic Analytics: If you want to go deeper into the data you have collected from users in order to understand "Why some things happened," you can use … intelligence tools to get some insights.

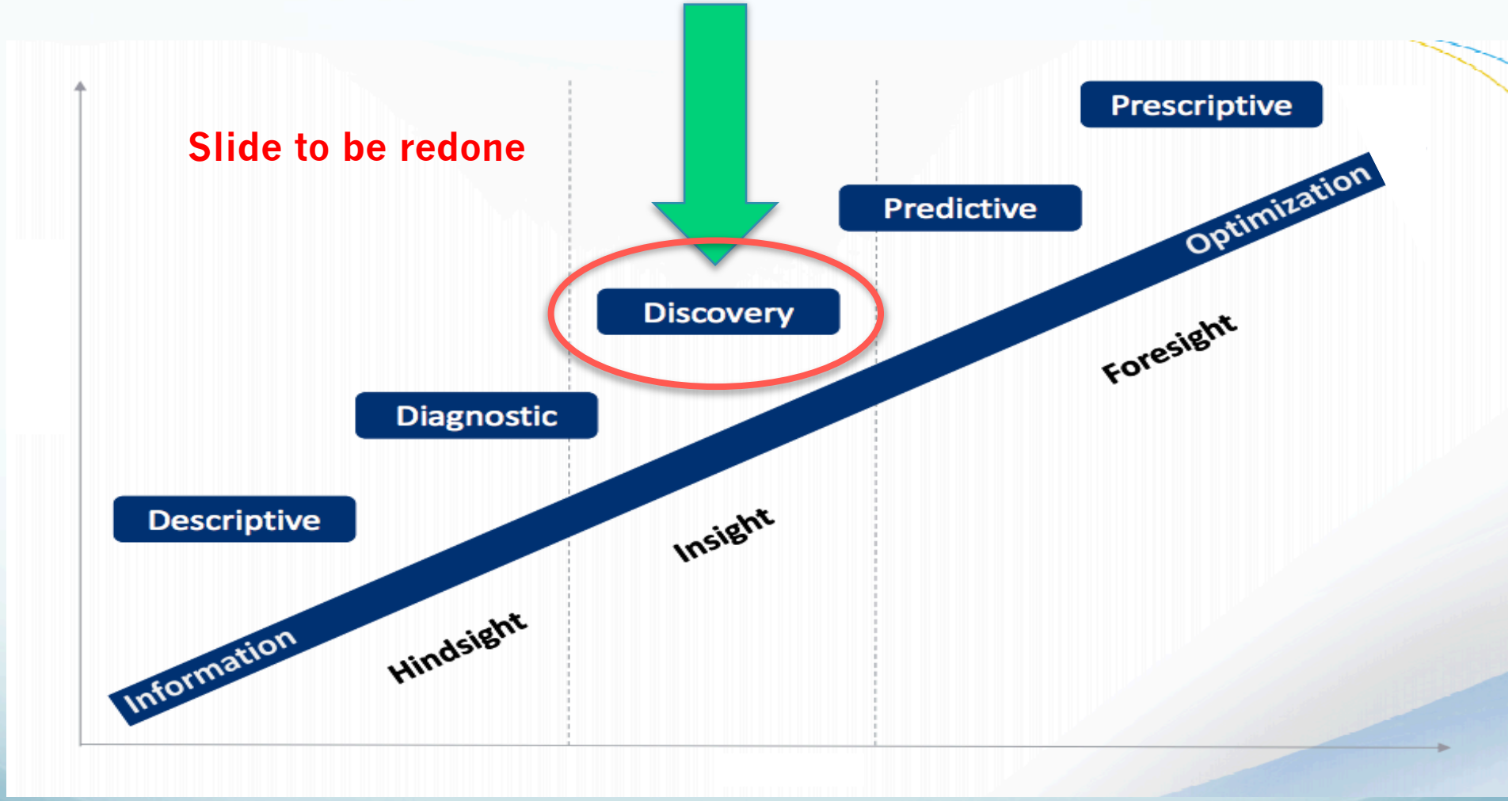
Diagnostic analytics
Determine *why* something happened, using content analytics and natural language processing to cull insights found in documents, email, websites, social media and so on. Understand the root cause of geophysical changes through more detailed analysis and visualizations.
(modified from: http://www.ibm.com/analytics/us/en/analytics-tools.html)

Diagnostic analytics looks deeper into what has happened and seeks to understand why a problem or event of interest occurs. How do various measurable events and actions in the focal domain relate to each other? (http://www.lifescaleanalytics.com/files/lifescale/files/brief_descriptivetoprescriptive.pdf)

Diagnostic data analytics is used to answer the question "Why is it happening?". It strives to identify root causes, key factors, and unseen patterns (
http://webcache.googleusercontent.com/search?q=cache:abyglyZBFLIJ:www.ag-ai.nl/download/17445-21-3-art.Parekh.pdf+&cd=8&hl=en&ct=clnk&gl=us&client=safari)

Discovery Analytics:

This is where people learn from the data.



http://www.informationbuilders.es/intl/co.uk/presentations/four_types_of_analytics.pdf

Discussion – Discoverative Analytics

Descriptive Analytics: You can quickly understand "what happened" during a given period in the past and verify if a campaign was successful or not based on simple parameters.

Diagnostic Analytics: If you want to go deeper into the data you have collected from users in order to understand "Why some things happened," you can use … intelligence tools to get some insights.

Discoveritive Analytics: The use of data and analysis tools/models to discover information

Predictive Analytics: If you can collect contextual data and correlate it with other user behavior datasets, as well as expand user data … you enter a whole new area where you can get real insights.

Prescriptive Analytics: Once you get to the point where you can consistently analyze your data to predict what's going to happen, you are very close to being able to understand what you should do in order to maximize good outcomes and also prevent potentially bad outcomes. This is on the edge of innovation today, but it's attainable!

Modified from: http://www.ciandt.com/card/four-types-of-analytics-and-cognition

Discussion – Discoverative Analytics

Discoveritive Analytics: The use of data and analysis tools/models to discover information

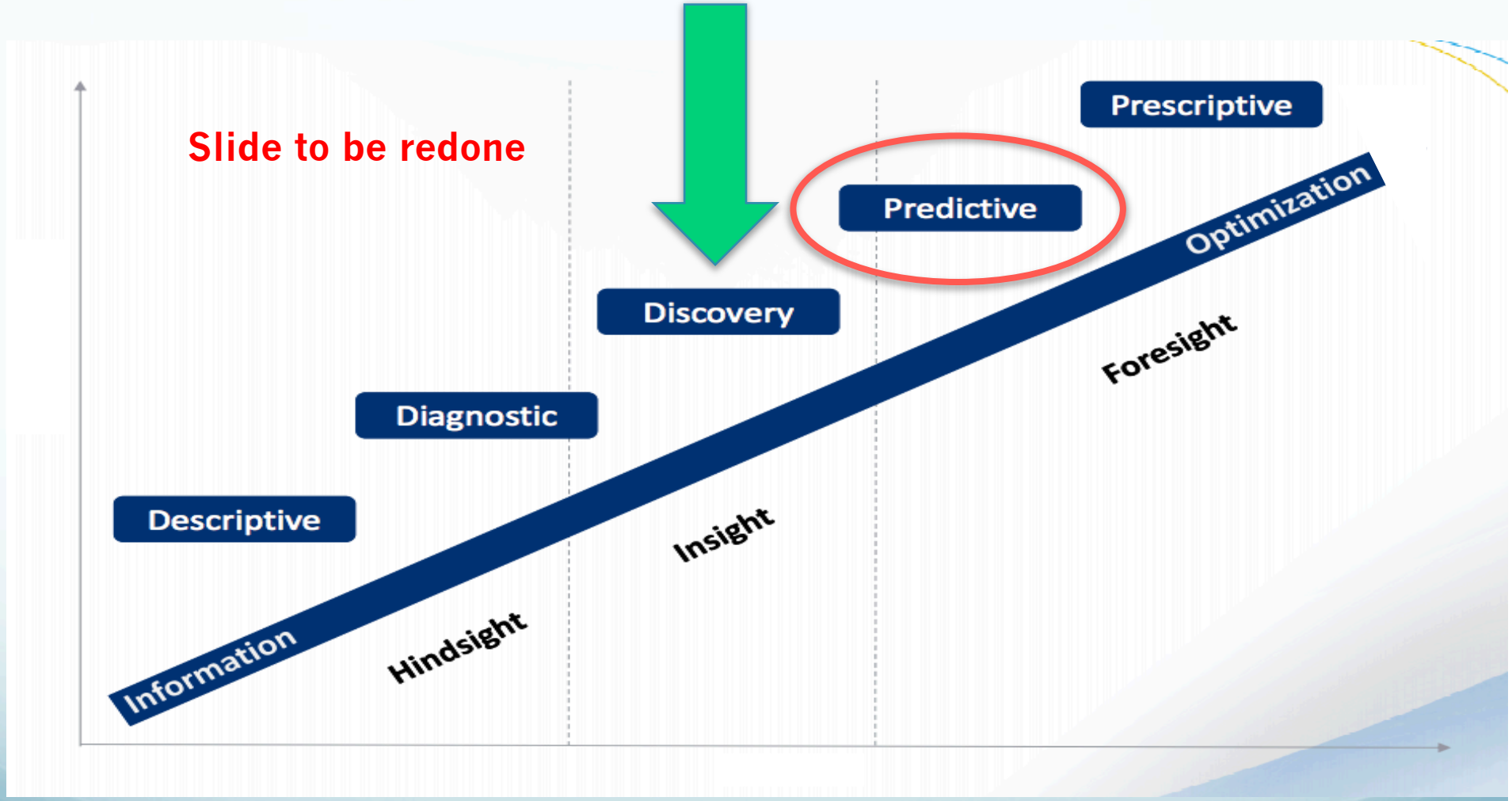
Discoverative Analytics

Tell me something that I don't know" is the definition of data mining - discovering unexpected patterns and relationships in data. (http://online-behavior.com/emetrics/data-discovery-1073)

Four types of discovery analytics: visual discovery, data discovery, information discovery and event discovery (http://www.information-management.com/blogs/3-major-trends-in-new-discovery-analytics-10024769-1.html)

Discovery Analytics:

This is where people learn from the data.



http://www.informationbuilders.es/intl/co.uk/presentations/four_types_of_analytics.pdf

Discussion – Predictive Analytics

Descriptive Analytics: You can quickly understand "what happened" during a given period in the past and verify if a campaign was successful or not based on simple parameters.

Diagnostic Analytics: If you want to go deeper into the data you have collected from users in order to understand "Why some things happened," you can use … intelligence tools to get some insights.

Discoveritive Analytics: The use of data and analysis tools/models to discover information

Predictive Analytics: If you can collect contextual data and correlate it with other user behavior datasets, as well as expand user data … you enter a whole new area where you can get real insights.

Prescriptive Analytics: Once you get to the point where you can consistently analyze your data to predict what's going to happen, you are very close to being able to understand what you should do in order to maximize good outcomes and also prevent potentially bad outcomes. This is on the edge of innovation today, but it's attainable!

Modified from: http://www.ciandt.com/card/four-types-of-analytics-and-cognition

Discussion – Predictive Analytics

Predictive Analytics: If you can collect contextual data and correlate it with other user behavior datasets, as well as expand user data … you enter a whole new area where you can get real insights.

Predictive Analytics

Encompasses a variety of statistical techniques from modeling, machine learning, and data mining that analyze current and historical facts to make predictions about future, or otherwise unknown, events

Combines techniques from statistics, data mining and machine learning to find meaning from large amounts of data…and predict where you're going.

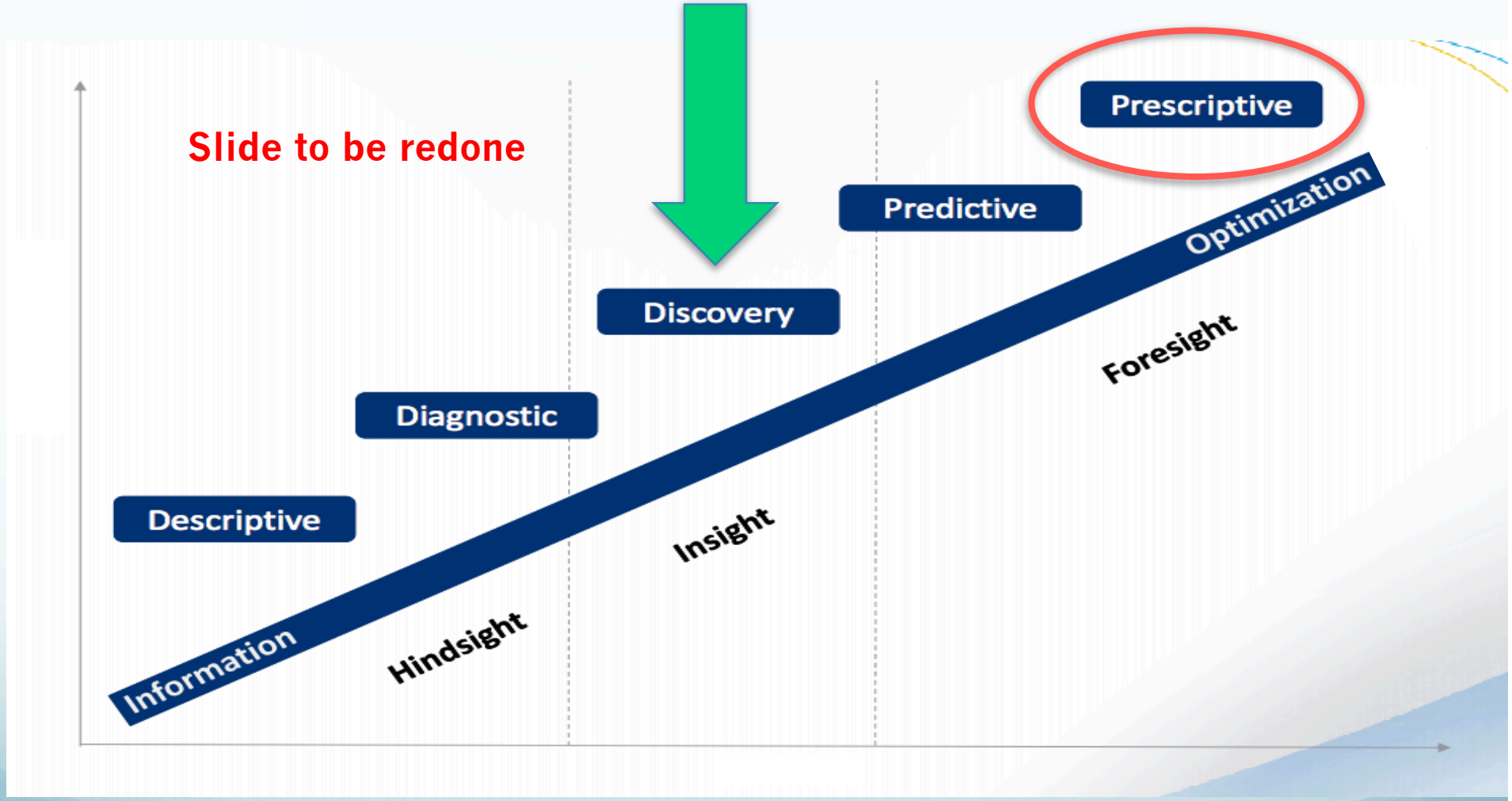
Predictive analytics is the practice of extracting information from existing data sets in order to determine patterns and predict future outcomes and trends. Predictive analytics does not tell you what will happen in the future. It forecasts what might happen in the future with an acceptable level of reliability, and includes what-if scenarios and risk assessment (http://www.webopedia.com/TERM/P/predictive_analytics.html)

Predictive analytics is the branch of data mining concerned with the prediction of future probabilities and trends. (http://searchcrm.techtarget.com/definition/predictive-analytics)

While regression analysis is commonly used, there exists another class of methods that deserve proper mentions. E.g. Bayes Network, Artificial Neural Net, Decision Tree, Support Vector Machine, etc. More importantly, the non-linear analysis aspect and the probability based approach that underpin many of the aforementioned methods.

Discovery Analytics:

This is where people learn from the data.



http://www.informationbuilders.es/intl/co.uk/presentations/four_types_of_analytics.pdf

Discussion – Prescriptive Analytics

Descriptive Analytics: You can quickly understand "what happened" during a given period in the past and verify if a campaign was successful or not based on simple parameters.

Diagnostic Analytics: If you want to go deeper into the data you have collected from users in order to understand "Why some things happened," you can use … intelligence tools to get some insights.

Discoveritive Analytics: The use of data and analysis tools/models to discover information

Predictive Analytics: If you can collect contextual data and correlate it with other user behavior datasets, as well as expand user data … you enter a whole new area where you can get real insights.

Prescriptive Analytics: Once you get to the point where you can consistently analyze your data to predict what's going to happen, you are very close to being able to understand what you should do in order to maximize good outcomes and also prevent potentially bad outcomes. This is on the edge of innovation today, but it's attainable!

Modified from: http://www.ciandt.com/card/four-types-of-analytics-and-cognition

Discussion – Prescriptive Analytics

Prescriptive Analytics: Once you get to the point where you can consistently analyze your data to predict what's going to happen, you are very close to being able to understand what you should do in order to maximize good outcomes and also prevent potentially bad outcomes. This is on the edge of innovation today, but it's attainable!

Prescriptive Analytics

Prescriptive analytics goes beyond descriptive and predictive models by recommending one or more courses of action -- and showing the likely outcome of each decision

Prescriptive analytics goes beyond predicting future outcomes by also suggesting actions to benefit from the predictions and showing the decision maker the implications of each decision option.

Descriptive vs. Diagnostic Analytics

Descriptive Analytics	Diagnostic Analytics
You can quickly understand "what happened" during a given period in the past and verify if a campaign was successful or not based on simple parameters.	If you want to go deeper into the data you have collected from users in order to understand "Why some things happened," you can use ... intelligence tools to get some insights
better understand what has happened	explain why an event happened
What has happened? How many? How much? Is this changing over time? The objective is to quantify, track and report what might have previously been only a vague qualitative sense for how things are going	looks deeper into what has happened and seeks to understand why a problem or event of interest occurs. How do various measurable events and actions in the focal domain relate to each other? While it may start with bivariate relationships it progress into development of multi-variable explanatory models
performed on historical data to establish statistical benchmarks in order to gain insights or answer the question ”What is happening?”	is used to answer the question “Why is it happening?”

Planning for January

Planning for January

Earth Science Data Analytics 101

Purpose: To 'educate' ESIP community on what Earth Science Data Analytics means, and provide exemplary use cases.

Cluster Goal: To stir innovation juices that can generate ideas/ techniques/collaborations/etc. that can facilitate/aid usage of data analytics

Draft Agenda

- Introduction to Earth science data analytics – (15 min)
- 3 or 4 use case speakers (10-15 min each) *I have 2 already…any suggestions*
- Current Data Analytics technologies useful in Earth science (15 min)
- Panel – Q&A (all speakers)

Planning for January

Earth Science Data Analytics 201

Purpose: To scope a study that would meaningfully benefit the ESIP and broad community; Develop an outline for the study

Cluster Goal: Publish our findings; Generate a library of Data Analytic methodologies

1. Take what we learn, refine, and define about the different types of Data Analytics

    - Descriptive Analytics
    - Diagnostic Analytics
    - Discoveritive Analytics
    - Predictive Analytics
    - Prescriptive Analytics

2. Associate exemplary Earth science use cases to each type

3. Associate Data Analytics techniques/tools to each type

4. Associate user categories to each type

5. Describe skills and expertise needed for each type

- Currently, we talk about our expertise and experience, but they seldom seem to connect to each other

- This will help us, the industry, and hopefully, educators, focus their understanding and interests regarding Earth Science Data Analytics.

Open Mic

User Model (Subsetted from ESDSWG WG)

Classes	Definition
Public	interested user of no or limited scientific skill
Graduate student	person of moderate to high skill at a university or college working towards an advanced degree
Production Centers	large organization that handles/processes vast quantities of data
Science Team	group of scientists focused on a specific area of study or on a specific instrument type, can include cal/val scientists
QA/Testing	developers or scientists using data to test software operation or to determine quality of a product, can include cal/val scientists
Data Analyst	person using NASA data to perform a specific analysis.
Domain Scientist	person using data to do research and publish within a discipline, comes in with some expertise in using the data
Interdisciplinary Scientist	person using high-level data products from multiple sources
Operational User	Data analyst or tech using data for operational support (applications) and emergency response
Assimilation Modelers	persons or groups that routinely obtain vast quantities of data for incorporation into models, can have operational needs

Relevant AGU Sessions

- Teaching Science Data Analytics Skills Needed to Facilitate Heterogeneous Data/Information Research: The Future Is Here - Session ID#: 1879
- Identifying and Better Understanding Data Science Activities, Experiences, Challenges, and Gaps Areas - Session ID#: 1809
- Advancing Analytics using Big Data Climate Information System - Session ID#: 3022
- Big Data in the Geosciences: New Analytics Methods and Parallel Algorithm - Session ID#: 3292
- Leveraging Enabling Technologies and Architectures to enable Data Intensive Science - Session ID#: 3041
- Open source solutions for analyzing big earth observation data - Session ID#: 3080
- Technology Trends for Big Science Data Management - Session ID#: 2525